

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 06/25/98	3. REPORT TYPE AND DATES COVERED Final Tech Report 12/1/92-11/30/96		
4. TITLE AND SUBTITLE Temporal processing with neural networks			5. FUNDING NUMBERS  N00014-93-1-0194	
6. AUTHOR(S)  Jeffrey L. Elman				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Regents of the Univ. of California University of California, San Diego 9500 Gilman Drive La Jolla, CA 92093			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research Code 1511: MIC Ballston Tower One 800 N. Quincy Street, Arlington, VA 22217-5660			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
<div style="border: 1px solid black; padding: 5px; display: inline-block;"> <b>DISTRIBUTION STATEMENT A</b>            Approved for public release;            Distribution is unlimited         </div> <div style="float: right; font-size: 2em; font-weight: bold; margin-top: 20px;">19980630 062</div>				
12a. DISTRIBUTION/AVAILABILITY STATEMENT  Unlimited				
13. ABSTRACT (Maximum 200 words)  <p>The research carried out under this contract focussed on four efforts, all involving the processing of temporal sequences by neural networks (1-3) or the effect of imposing a spatio-temporal gradient on network learning (4):</p> <p>(1) Assessing alternative neural network techniques for problems involving temporal coding.</p> <p>(2) Development of tools for analysing recurrent networks, so that the solutions of successfully trained networks can be better understood;</p> <p>(3) Development of a dynamical systems theory approach to computation in recurrent networks.</p> <p>(4) Development of biologically and cognitively plausible techniques for enhancing training.</p>				
14. SUBJECT TERMS			15. NUMBER OF PAGES 10	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT -	



CENTER FOR RESEARCH IN LANGUAGE

9500 GILMAN DRIVE, 0526  
LA JOLLA, CALIFORNIA 92093-0526

June 25, 1998

Office of Naval Research  
San Diego Regional Office  
4520 Executive Drive, Suite 300  
San Diego, CA 92121-3019

Attn: Ms. Veronica Lacey

Reference: Agreement No. N00014-93-1-0194

As required under the terms and conditions of the referenced agreement,  
enclosed is a copy of Form 298 and the final technical report.

Sincerely,

A handwritten signature in cursive script that reads "Jan M. Corte".

Jan M. Corte  
Center for Research in Language

Enclosures: 2

cc: Scientific Officer, Code 1142PS (3)  
Director, NRL, code 2627 (1)  
DTIC (2)  
OCGA, UCSD (1)

# **FINAL REPORT**

**Contract ONR N00014-93-0194**

**Project title:** Temporal processing with neural networks

**Principal investigator:** Jeffrey L. Elman

**Period of time covered by report:** 12/1/92-11/30/96

## **1. Brief overview of scientific achievements:**

The research carried out under this contract focussed on four efforts, all involving the processing of temporal sequences by neural networks (1-3) or the effect of imposing a spatio-temporal gradient on network learning (4):

(1) Assessing alternative neural network techniques for problems involving temporal coding.

(2) Development of tools for analyzing recurrent networks, so that the solutions of successfully trained networks can be better understood;

(3) Development of a dynamical systems theory approach to computation in recurrent networks.

(4) Development of biologically and cognitively plausible techniques for enhancing training.

Work in the initial area was carried out with Thomas Rebotier (a doctoral student in the Cognitive Science department at UCSD). We constructed a suite of problems, including economic indices, seismic data, strings generated from various classes of grammars, speech data, and acoustic data. Second, we have developed local implementations of Time Delay Neural Networks, Hidden Markov Models, Backpropagation Through Time, and Simple Recurrent Networks. [This work was reported in detail in earlier progress reports and will not be summarized here.]

Work in the second and third areas was done in collaboration with Rebotier, Paul Rodriguez (another doctoral student in Cognitive Science) and Janet Wiles (University of Queensland). We used various techniques for analyzing the movement of networks' internal state vectors through state space, over time; the goal is to understand how the networks make use of state space and temporal dynamics to encode temporal information.

Work in the final section was done with Thomas Rebotier, in collaboration with Mark Johnson (University College of London) and Jeff Shrager (University of Pittsburgh). Development of biologically and cognitively plausible techniques for enhancing training. The goals were two-fold: (i) to account for the spatial differentiation, over time, of initially multipotent embryonic cortex; and (ii) to extend the computational capacities of Hebbian learning by imposing a spatio-temporal gradient on the learning process.

## **2. Summary of results**

### **2.1 Benchmarks.**

This work was reported in detail in earlier progress reports and will not be summarized here.

### **2.2 Analytic tools.**

A common complaint regarding neural networks (when they are offered as models of biological processes) is "what value is there in replacing one black box (e.g., the brain) with another black box (e.g., a neural network that emulates some brain-like capability)." The complaint is reasonable, since much of the value of a model presumably lies in our ability to probe it to a greater extent than is possible with biological systems (for which invasive procedures are limited and highly regulated, and for which non-invasive tests have a lesser degree of precision). However, it is also a complaint which is now somewhat dated, since most network researchers recognize that network analysis plays a crucial role in validating the models they construct.

Under the current contract, a number of novel network analyses have been developed and utilized. Many of these are not novel in other fields, but their application to neural network research is (and in many cases, the research supported by the current contract was the first to utilize them). These include the use of principal components analysis; projection pursuit; multidimensional scaling; and contribution analysis. In the relatively short period since the beginning of this contract (1992), many of these tools have become standard in the field.

### **2.3 Development of a dynamical systems account of computation in recurrent neural networks.**

Relatively little is still known about the computational properties of recurrent networks. Despite early proofs about Turing capability (J. Pollack's thesis), and more recent important work by Seligmann and Sontag, we still lack the kind of analysis for recurrent networks which the Chomsky hierarchy provides for discrete automata. The Chomsky hierarchy maps machines onto grammars, and indicates the computational benefits which result from extending machine resources in a principled way.

Recently, as part of the work supported by this contract, Janet Wiles, Paul Rodriguez, and I have undertaken a series of studies which have as their goal understanding how traditional formal languages (as classified by the Chomsky hierarchy) might be processed by recurrent networks. Ultimately, we hope to understand what the natural classes of languages are (because conceivably the natural class of languages processed by recurrent networks may not be commensurate with the class of languages defined under the Chomsky hierarchy). In the short term, our focus is to understand how recurrent networks carry out computation; our specific perspective has been to study this using dynamical systems analysis.

In initial studies, we found that a recurrent network could process a Context Free Grammar ( $a^n b^n$  e.g., some number of  $a$ 's followed by an equal number of  $b$ 's) by setting two dynamical regimes. In the first regime (when the network is in "counting up" mode, inputting  $a$ 's), the network has an attracting fixed point (with oscillatory behavior); this is shown in Figure 1a. In the second regime (when the network is in

“counting down” mode, receiving  $b$ 's), the network has a repelling fixed point (again, with oscillatory behavior); this is shown in Figure 1b. By precisely equilibrating the rate of contraction of the attracting fixed point with the rate of expansion of the repelling fixed point, and by ensuring that in the transition from the last  $a$  to the first  $b$ , the network moves to a distance from the second fixed point which is matched to the distance from the first fixed point, the network guarantees that when the final  $b$  is input it will recognize the end of string. In more recent studies, we have extended this work to more complex languages, including the palindrome language  $xx^R$  (a string of inputs,  $x$ , followed by  $x^R$ , the string in reversed form). This language is interesting because it resembles center-embedded relative clauses found in natural language (e.g., “The book that the girl read is missing.”).

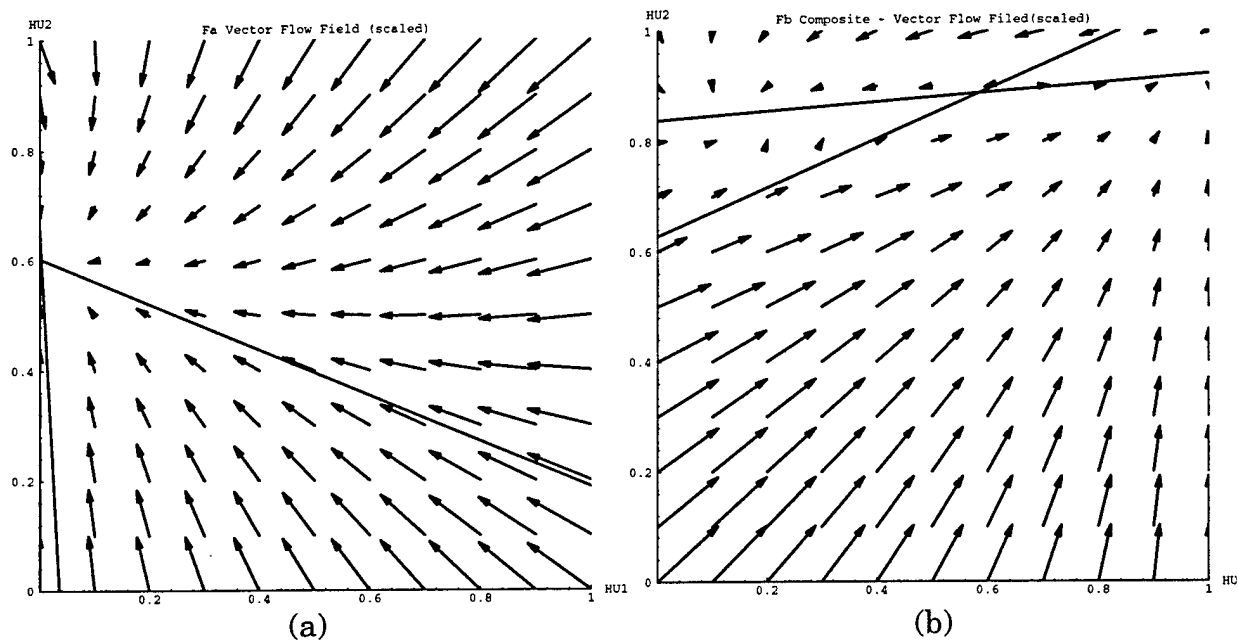


Figure 1. Vector flow fields for two dynamical regimes of network trained on  $a^n b^n$  language. (a) Flow field while  $a$ 's are input; (b) flow field while  $b$ 's are input.

## 2.4 Interactions between learning and timing.

Many of the neural models for temporal processing which have been studied to date suffer from scaling problems. The models work well with restricted data or on toy problems. Attempts to scale up to larger data sets or to time series in which the temporal relationships are more complex often do not work well. This problem of scaling is of course not unique to neural network models; the failure to scale is a chronic problem of many models.

I have recently become interested in the possibility that the developmental trajectory which humans undergo may interact with the learning of complex behaviors. An inordinately long portion of the human life cycle is spent in a period of

immaturity; given the vulnerability of the immature state this would seem to be evolutionarily maladaptive. On the other hand, there may be positive consequences to delayed development. My hypothesis is that in fact certain problems are best learned by "starting small"—i.e., with limited resources. I have been studying this possibility in two realms.

(a) In the first, I attempted to train a simple recurrent network to process strings generated by a context-free grammar. (This is a category of formal languages into which human languages are minimally classified; human languages may in fact be somewhat more complex.) Although humans appear able to learn such grammars, recurrent networks consistently failed, across a wide ranging of training conditions. However, when the networks were trained in an incremental fashion, they succeeded in learning the data sets. Incremental training was carried out in either of two ways (both worked equally well). In the first regime, networks were trained on a subset of strings which were shorter in duration and which contained no embeddings. After mastering this simpler data set, the networks were given increasingly more complex data. In the second regime, networks were trained from the beginning on the final complex data set. However, noise was injected every two or three tokens during early portions of training. This noise effectively interfered with the learning of the more complex data. As training progressed, the periodicity of the noise was increased in two or three word increments and eventually eliminated. In both regimes, learning was rapid and generalization was high. The technique is similar to a hypothesis proposed by Newport for children. The assumption is that early resource limitations force the networks to focus on the major sources of variance, and that this provides a necessary scaffolding for the networks to learn more complex interactions exhibited by longer sentences. Thus, for the networks to achieve the final "adult" competence requires that they go through a maturational period which resembles that of children.

Figure 2a shows the internal state space of a network which was trained in the non-incremental fashion; the state space is relatively unstructured and fails to encode temporally significant information. Figure 2b shows the internal state space of a network which was trained in the incremental manner described above. The state space is well-structured and encodes grammatically relevant information.

(b) In the second series of experiments (done in collaboration with Jeff Shrager and Mark Johnson) I have been interested in the possible computational benefits of another developmental pattern, namely the fact that neo-natal human cortex undergoes waves of synaptic proliferation followed by synaptic pruning. These waves do not occur everywhere simultaneously. Instead, they last over a period of several years and pass over different spatial regions of cortex at different points in time.

The initial series of experiment, carried out by Kerszberg, Dehaene, and Changeux and replicated by Shrager, Johnson, and myself, involved a slab of 'pseudo-cortex', shown in Figure 3. This slab consisted of a 30x30 matrix of nodes. Each node received random inputs from neighbors following a Gaussian probability distribution, such that connections from near neighbors was more probable than from distal units. Each node in addition received input from two afferents (marked A

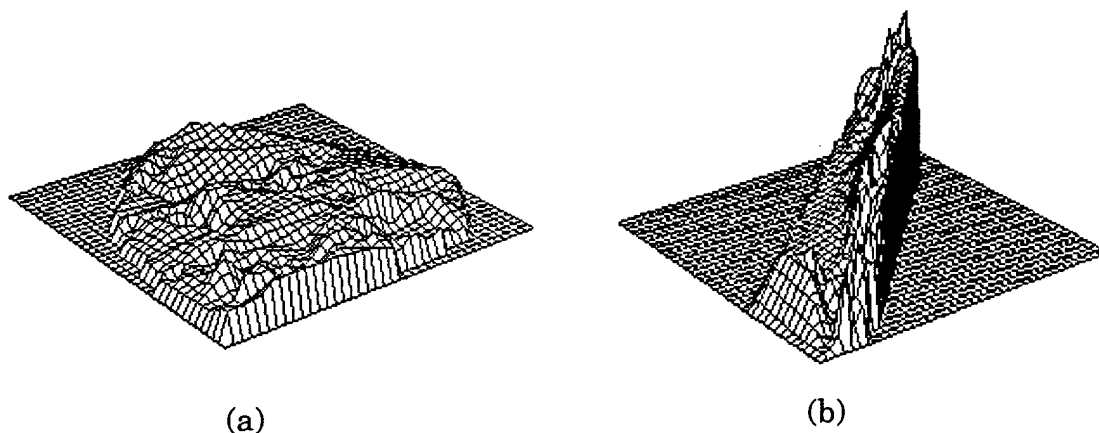


Figure 2. Plot of hidden unit activation patterns (in response to presentation of 10,000 word inputs), shown in coordinates of first 3 principal components. (a) Network which failed in the task. (b) Network which succeeded in the task, after being trained with an incremental regime.

and B), which fired randomly and simultaneously. The connection matrix for the slab was modified according to a Hebbian learning rule. After learning, the question was asked of each node, what function of the two afferents is it computing. Most nodes remained off, but some fired whenever A was on; others fired whenever B was on; others became AND units, etc. This was the expected result.

However, when learning progresses in a staged manner—modeling the movement over time of a “trophic factor” through the matrix, such that columns underneath the TF wave are more plastic while those elsewhere decay or do not learn—then a different result is obtained. We have found that if the TF wave moves from left to right, so that the left-most columns are early learners and the right-most columns are late learners, then the units on the left develop as in the first condition. However, a significant number of units in the late learning columns become XOR units. This is a surprising result given the known problem with Hebbian learning and non-correlated input patterns; it results from the fact that the TF wave allows early learning units to develop which become sensitive to OR and AND functions. The late learning units then take as their input not only the external afferents but the outputs of these OR and AND units, and that makes it possible for them to learn XOR. This result is promising because it demonstrates that a learning rule of known biological plausibility but known computational limitation may be “salvaged” by subjecting learning to a maturational regime which is itself plausible.

The work to date only involves temporally static stimuli. We are now trying to extend this finding to conditions in which stimuli are more complex and which involve temporal dependencies.

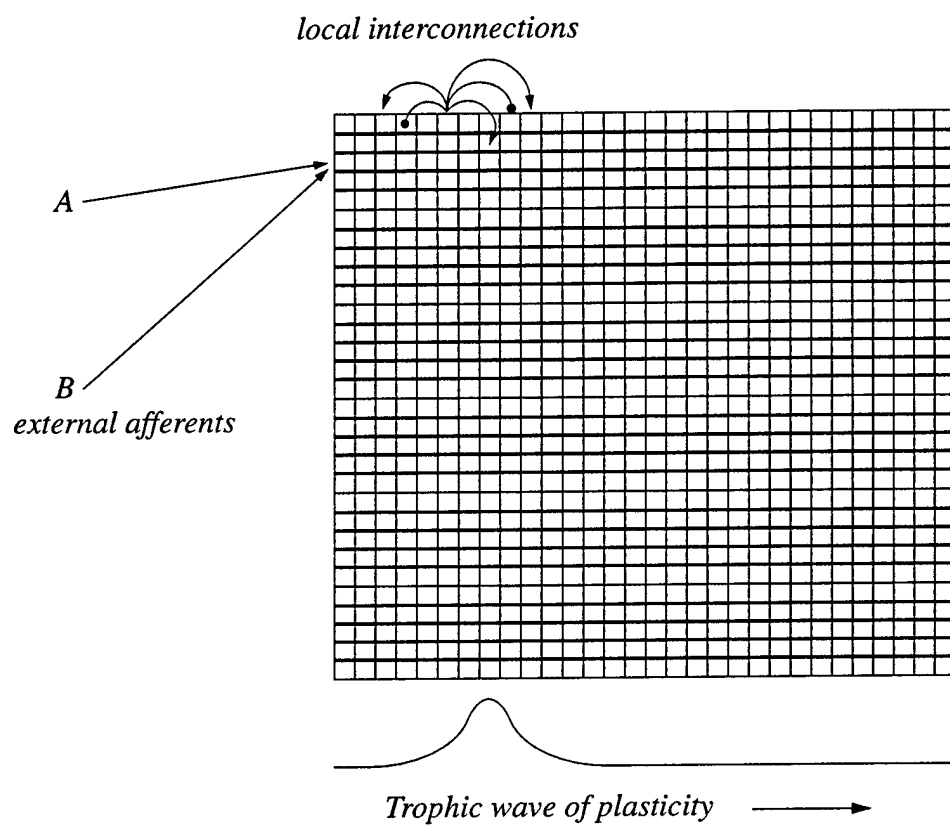


Figure 3



### 3. Productivity report:

#### (a) Books:

- Mozer, M., Smolensky, P., Touretzky, D., Elman, J., & Weigand, A. (1994). *Proceedings of the 1993 Connectionist Models Summer School*. Hillsdale, NJ: Erlbaum.
- Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Plunkett, K., & Elman, J.L. (in press, 1997). *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*. Cambridge, MA: MIT Press.

#### (b) Journal articles:

- Elman, J. L. (1992). Grammatical structure and distributed representations. In Steven Davis (Ed.), *Connectionism: Theory and Practice*. Oxford: Oxford University Press.
- Hare, M., & Elman, J.L. (1992). A connectionist account of English inflectional morphology: Evidence from language change. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Weckerly, J., & Elman, J.L. (1992). A PDP approach to processing center-embedded sentences. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71-99.
- Bates, E., & Elman, J. L. (1993). Connectionism and the study of change. In Mark Johnson (Ed.), *Brain development and cognition: A reader*. Oxford: Blackwell Publishers.
- Hare, M., & Elman, J.L. (1993). From weared to wore: A connectionist account of language change. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum. Pp. 528-533.
- Shultz, T.R., & Elman, J.L. (1994). Analyzing cross-connected networks. In J.D. Cowan, G. Tesauero, and J. Alspector (Eds.), *Advances in Neural Information Processing Systems 6*. San Mateo, CA: Morgan Kaufmann.
- Elman, J. L. (1994). Learning and development in neural networks: The importance of starting small. In C. Umiltà and M. Moscovitch (Eds.), *Attention and Performance XV: Conscious and Nonconscious Information Processing*. Hillsdale, NJ: Erlbaum.
- Bates, E., Elman, J.L. & Li, P. (1994). Language in, on and about time. In M. M. Haith, J.B. Benson, R. J. Roberts, and B. F. Pennington (Eds.), *The Develop-*

*ment of Future-Oriented Processes*. Chicago: University of Chicago Press.

- Thyme, A., Ackerman, F., & Elman, J.L. (1994). Finnish nominal inflections: Paradigmatic patterns and token analogy. In S.D. Lima, R.L. Corrigan, and G.K. Iverson (Eds.), *The reality of linguistic rules*. John Benjamins. Pp. 445-466.
- Nolfi, S., Elman, J.L., & Parisi, D. (1994). Learning and evolution in neural networks. *Adaptive Behavior*, 3:1, 5-28.
- Hare, M., & Elman, J.L. (1995). Learning and morphological change. *Cognition*, 56, 61-98.
- Hare, M., Elman, J.L., & Daugherty, K.G. (1995). Default categorization in connectionist networks. *Language and Cognitive Processes*, 10, 601-630.
- Elman, J.L. (1995). Language processing. In M. Arbib (Ed.) *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press. Pp. 508-513.
- Elman, J.L. (1995). Language as a dynamical system. In R. Port and T. van Gelder (Eds.), *Mind as Motion: Dynamical Perspectives on Behavior and Cognition*. Cambridge, MA: MIT Press.
- Rebotier, T., & Elman, J.L. (1995). Explorations with the dynamic wave model. In M. Mozer (Ed.), *Advances in Neural Information Processing Systems 7*. San Mateo, CA: Morgan Kaufmann.
- Wiles, J., Rodriguez, P., & Elman, J.L. (1995). Learning to count without a counter: A case study of dynamics and activation landscapes in recurrent networks. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Bates, E., & Elman, J. (1996). Learning rediscovered: A perspective on Saffran, Aslin, and Newport. *Science*, 274:5294 (December 13, 1996.), 1849-1850.
- Plunkett, K., Karmiloff-Smith, A., Bates, E.A., Elman, J.L., & Johnson, M.H. (1997). Connectionism and developmental psychology. *Journal of Child Psychology and Psychiatry*, 38, 53-80.
- Elman, J.L., & Hare, M.L. (1997). Single mechanism= single representation? No!! In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Bates, E., Elman, J., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (in press). On innateness. In William Bechtel & George Graham (Eds.), *A Companion to Cognitive Science*. Oxford: Basil Blackwood.
- Elman, J.L. (in press). Connectionism, artificial life, and dynamical systems. In William Bechtel & George Graham (Eds.), *A Companion to Cognitive Science*. Oxford: Basil Blackwood.
- Elman, J.L. (in press). The emergence of language: A conspiracy theory. In B. MacWhinney (Ed.), *The emergence of language*. Hillsdale, NJ: Lawrence Earl-

baum Associates.

(c) *Invited presentations:*

April, 1993: Duke University: Workshop on Temporal Coding in Biological Systems. Invited talk.

May, 1993: University of Wisconsin: Society for Research in Child Language Disorders. Keynote Address.

October, 1993: University of California, Irvine. Cognitive Science Department. Invited talk.

February, 1994: Washington University, St. Louis: Program in Philosophy, Neuroscience, and Psychology. Invited talk.

March, 1994: Ohio State University: Cognitive Science Program. Invited talk.

April, 1994: Rice University: Cognitive Science Program. Invited talk. Linguistics Department. Invited talk.

April, 1994: California State University, Fullerton. Cognitive Science Program. Invited talk.

April, 1994: City University of New York: Linguistics, Cognitive Science, and Childhood Language Disorders. Tutorial and invited talk.

December, 1994: University of San Marino, San Marino. Symposium on Rethinking Innateness. Invited talk.

April, 1996: International Conference on Infant Studies. Invited talk.

May, 1996: Santa Fe Institute: Workshop on Dynamical Models of Cognition Invited talk.

May, 1996: Society for Philosophy and Psychology. Invited talk.

September, 1996: Sussex University, England. Invited talk.

September, 1996: University of York, England. Invited talk.

October, 1996: University of Oxford, England. McDonnell-Pew Center for Cognitive Neuroscience. Invited talk.

October, 1996: University of Colorado at Boulder. Invited talk.

November, 1996: University of Texas at Austin. Invited talk.

February, 1997: University of Chicago. Invited talk.

March, 1997: UCSD/Salk Center for Cognitive Neuroscience Retreat. Invited talk.

March, 1997: 10th Annual CUNY Conference on Sentence Processing (Los Angeles). Invited talk.

May, 1997: Carnegie Symposium, Carnegie Mellon University. Invited talk.  
August, 1997: Computational Psycholinguistics Conference (Stanford). Invited talk.  
August, 1997: University of Texas at Austin. Department of Psychology. Invited talk.  
August, 1997: Computational Psycholinguistics Workshop (Berkeley). Invited talk.  
November, 1997: McGill University. Cognitive Science Program. Hebb Distinguished Speaker Address.  
February, 1998: Cornell University Cognitive Science Program. Invited talk.  
May 4: University of California, Irvine. Department of Cognitive Science. Invited talk.  
May 19: Hunter College, CCNY. Department of Psychology. Invited talk.

*(d) student training:*

graduate students:

Arshavir Blackwell (Psychology/Cognitive Science, UCSD)  
Jay Moody (Cognitive Science, UCSD)  
Thomas Rebotier (Cognitive Science, UCSD)  
Paul Rodriguez (Cognitive Science, UCSD)  
Jill Weckerly (Cognitive Science/Linguistics, UCSD)

postgraduate scholars/visitors:

Mary Hare (CRL, UCSD)  
Thomas Shultz (Psychology, McGill Univ.)  
Janet Wiles (Computer Science, Univ. of Queensland)